Research Report

The use of new data sources in small area estimation of social cohesion

Lidiya Mishieva (2975882)
Supervisors: Angelo Moretti, Camilla Salvatore and Peter Lugtig

19 January, 2026

Table of contents

List of Tables

1 Introduction

In the political, public and academic debate, social cohesion is commonly regarded as a feature that always seems to be currently deteriorating. The origins of the concept of social cohesion can be traced back to classical sociological research, which sought to understand the ‘social glue’ holding modern societies together (Durkheim 2018; Ferdinand et al. 1887). Contemporary research on social cohesion has largely moved away from general sociological theory, but has not arrived at a universally accepted definition yet (Schiefer and Noll 2016). Policy-oriented research mainly focuses on measurement models of social cohesion, where social cohesion is typically defined as a latent, not directly measurable, property of populations (Chan, To, and Chan 2006), consisting of multiple dimensions (Jenson (1998); Duhaime et al. (2004); Berger-Schmitt (n.d.); Noll (2002); Rajulton, Ravanera, and Beaujot (2006); Bollen and Hoyle (1990); Whelan and Maître (2005); Chan, To, and Chan (2006)).

Bottoni (2016) proposes and validates a latent multi-level model of social cohesion as a synthesis of multiple contributions in the literature. Their model distinguishes between the collective and individual dimension of the phenomenon, where social cohesion is treated as a collective property of a population, whereas social inclusion is conceptualized as an individual-level attribute. Dimensions of social cohesion used in Bottoni’s model are relational in nature, capturing both practical interactions between individuals (*social support*, *participation*, *density of relations*) and attitudinal dispositions towards individuals, institutions or groups (*interpersonal* and *institutional trust*, *openness towards out-groups*, *legitimacy of institutions*).

Designing targeted policies to strengthen social cohesion requires evidence on how cohesion varies across sub-national areas. While some dimensions, such as trust, civic participation, and attitudes toward immigration, have been shown to vary spatially within countries (Brakel, Boonstra, and Smeets 2025); Moretti and Ahmed (2024); Czaika and Di Lillo (2018); Orteca (2015)], systematic evidence on the small-scale spatial distribution of social cohesion across Europe remains limited.

Being measured at individual level, social cohesion indicators need to be aggregated to regional level. However, survey data is typically not designed for obtaining efficient estimates at a sub-national level. Small area estimation (SAE) offers a methodological tool capable of producing more accurate and precise estimates using survey data. SAE consists of estimating target parameters for groups or ‘areas’ where the sample size is small or which do not appear in the sampling design using statistical models. This is done by ‘borrowing strength’ via incorporating auxiliary information correlated with target estimates (Rao and Molina 2015). Compared to direct estimates, SAE typically lowers the mean squared error by leveraging the bias-variance trade-off. For social cohesion, multivariate SAE (Benavent and Morales 2016) could help making use of the relatedness of different social cohesion dimensions and improve the estimated even further.

Geographic big data can potentially help explaining spatial variability in outcomes modeled through small area estimation by offering high-resolution geo-referenced data of built and natural environment. OpenStreetMap (OSM) represents one of the most prominent of these sources (Goodchild 2013). Although some studies highlight its potential for official statistics Ninivaggi and Cutrini (2023), its use as auxiliary information in SAE remains unexplored. OSM's data quality and completeness have been subject of extensive investigation. In terms of coverage, regions with higher building density are better covered, meaning that urban areas better represented in OSM (Zhou 2017). But there are differences: urban areas which are located in Europe, are economically better off and more urbanized have the highest building completeness (Herfort et al. 2023). While street network is well covered across the globe (Barrington-Leigh and Millard-Ball 2017) feature-dependent completeness for buildings was evaluated only in some local studies (Balducci 2019; Maina et al. 2019).

Against this background, the question arises: To what extent can small area estimates of social cohesion at the sub-national level in Europe be improved by the use of new data sources? This thesis is organized as follows. Section 2 reviews the literature on the immediate residential environment as a determinant of social cohesion, with the aim of identifying suitable covariates for the small-area estimation models that can be derived from OSM. Section 3 provides an overview of the data sources and variables used in the analysis. Section 4 then describes the methodological approach.

2 Theoretical background

Relations between individuals

Spatial composition and spatial configuration are characteristics of physical space, shaping the conditions under which social interaction between individuals becomes more or less likely (Small and Adler 2019). Spatial composition refers to the physical elements that constitute a space and can serve as locations for social interaction. However, not all places contribute equally to social relations. Places serving as intentional gathering places can act as foci around which people organize their interactions and thereby will form interpersonal ties and clusters of social connections (Feld 1981). By contrast, spaces that bring people together only through mere co-presence, without supporting shared activities or a sense of collective identity, may have limited impact on building social cohesion among residents (Wickes et al. 2018).

Spatial configuration refers to the arrangement of land use and the segmentation of physical space by man-made or natural boundaries, which can act as barriers to social interaction (Small and Adler 2019). The extent to which its layout allows social ties to move easily across space can impose constraints on how social relations are built and maintained within this space (Hipp et al. 2014). Low-density, single-use layouts in residential environments reduce perceived safety for pedestrians and cyclists and can discourage movement (Cho, Rodríguez, and Khattak 2009). Individuals perceive walking in parks as safer and more attractive than walking through residential areas (Basu et al. 2022). In addition, some spatial features can have ambivalent effects: retail destinations may attract outsiders whom local residents interpret as a potential threat to safety (Foster et al. 2013).

Relations between individuals and groups

Social infrastructure – spaces such as libraries, parks, educational institutions, and community centers – is commonly understood as infrastructure that serves as networking space thereby facilitating inter-personal trust (Joshi and Aldrich 2022). These interactions can also facilitate positive relations between different social groups (Christie and Allport 1954). Areas with greater availability of social infrastructure expose residents to a broader range of potentially heterogeneous social contacts (Heine et al. 2025; Fraser, Awadalla, et al. 2024). Persistent inequalities in access to amenities driven by social segregation across space (Michelangeli et al. 2025) may limit the potential for cross-group encounters. Further, social segregation is reproduced within certain types of social infrastructure by reinforcing interactions among similar groups rather than fostering diversity (Fraser, Yabe, et al. 2024).

Relations between individuals and institutions

The quality of local infrastructure shapes how individuals evaluate and relate to public institutions. When residents feel satisfied with the amenities and services in their residential area, they tend to express higher trust in local authorities, which can serve as a foundation to general public trust (Fitzgerald and Wolak 2014). At the same time, perceptions of deteriorating or unequal service provision can erode institutional trust, as seen in widening urban–rural divides where rural residents increasingly view their socio-economic infrastructure as lagging behind that of urban centers (McKay, Jennings, and Stoker 2023; Mitsch, Lee, and Ralph Morrow 2021). Dissatisfaction with local amenities can prompt people to move elsewhere (Dustmann and Okatenko 2014; Ulrich-Schad 2015), suggesting that institutional performance additionally indirectly shapes trust by influencing residential stability and the ability to sustain long-term community ties.

3 Data

3.1 Survey data

For measuring social cohesion, the operationalization is based on the latent model proposed by Bottoni (2016). The model was originally validated using indicators from the European Social Survey (ESS) Round 6 (2012). The ESS is a biennial, cross-national survey based ont cross-sectional probability sampling of all persons aged 15 years and older residing in private households within each participating country. In ESS, the data is geo-referenced through Nomenclature of Territorial Units for Statistics (NUTS) regions. The available levels of aggregation of geographic location is different across European countries as illustrated in Table 1.

Table 1: Available NUTS levels for respondents' place of residence by country in ESS 11

NUTS level	ESS country
NUTS 1	Cyprus, Germany, Italy, Montenegro, United Kingdom
NUTS 2	Austria, Belgium, France, Greece, Latvia, Netherlands, Norway, Poland, Portugal, Serbia, Spain, Sweden, Switzerland
NUTS 3	Bulgaria, Croatia, Finland, Hungary, Iceland, Ireland, Lithuania, Slovakia, Slovenia

Bottoni's original framework includes seven dimensions of social cohesion, but this thesis models only five due to data limitations in ESS Round 11 (2023). The dimensions of social support and participation are excluded because the corresponding indicators are unavailable. Using earlier ESS rounds would allow their inclusion but would require outdated and incomplete OSM data. The selected approach prioritizes consistency with current, higher-quality OSM information. The dimensions and the indicators are displayed in Table 2.

Table 2: ESS round 11 questions measuring five dimensions of social cohesion

Dimension	Indicator	Measurement
1. Interpersonal trust	1. Most people can be trusted or you can't be too careful	0=*You can't be too careful* 10=*Most people can be trusted*
	2. Most people try to take advantage of you, or try to be fair	0=*Most people try to take advantage of me* 10=*Most people try to be fair*
	3. Most of the time people helpful or mostly looking out for themselves	0=*People mostly look out for themselves* 10=*People mostly try to be helpful*

Dimension	Indicator	Measurement
2. Density of social relations	4. How often socially meet with friends, relatives or colleagues	1=*Never* 2=*Less than once a month* 3=*Once a month* 4=*Several times a month* 5=*Once a week* 6=*Several times a week* 7=*Every day*
	5. How many people with whom you can discuss intimate and personal matters	0=*None* 1=*1* 2=*2* 3=*3* 4=*4-6* 5=*7-9* 6=*10 or more*
	6. Take part in social activities compared to others of same age	1=*Much less than most* 2=*Less than most* 3=*About the same* 4=*More than most* 5=*Much more than most*
3. Openness	7. Immigration bad or good for country's economy	0=*Bad for the economy* 10=*Good for the economy*
	8. Country's cultural life undermined or enriched by immigrants	0=*Cultural lift undermined* 10=*Cultural life enriched*
	9. Immigrants make country worse or better place to live	0=*Worse place to live* 10=*Better place to live*
4. Institutional trust	10. Trust in country's parliament	0=*No trust at all* 10=*Complete trust*
	11. Trust in the legal system	0=*No trust at all* 10=*Complete trust*

Dimension	Indicator	Measurement
	12. Trust in politicians	0=*No trust at all* 10=*Complete trust*
	13. Trust in political parties	0=*No trust at all* 10=*Complete trust*
5. Legitimacy of institutions	14. How satisfied with the national government	0=*Extremely dissatisfied* 10=*Extremely satisfied*
	15. How satisfied with the way democracy works in country	0=*Extremely dissatisfied* 10=*Extremely satisfied*
	16. State of education in country nowadays	0=*Extremely bad* 10=*Extremely good*
	17. State of health services in country nowadays	0=*Extremely bad* 10=*Extremely good*

3.2 Auxiliary data

3.2.1 Adminstrative data

For administrative data, we use statistics from Eurostat for EU countries and from NOMIS for the UK. These sources provide reliable and comparable administrative data for NUTS-regions. The variables included in the analysis are population density, GDP per capita, employment, and number of police-recorded offences.

3.2.2 OpenStreetMap data

OSM data can be accessed via the read-only application programming interface Overpass API using the Overpass Query Language (Overpass QL), which supports filtering by attributes, geometry, and geographic extent. The database is structured around three core elements: nodes (point coordinates), ways (ordered lists of nodes forming lines or areas), and relations (groupings of nodes and ways). Attributes in OSM are described using tags, which are flexible key–value pairs attached to nodes, ways, and relations.

An OSM snapshot dated 01.01.2023 was used to match ESS11, with queries limited to the administrative boundaries of ESS11 countries that use NUTS classification for official statistics. All attributes documented for each feature were extracted. The key-value pairs used in the queries are listed in Table 3. The choice of the features is based on the literature review introduced in Chapter 2.

Table 3: Key-value pairs of OpenStreetMap features used for Overpass API queries

Type of infrastructure	Key	Value
Buildings		
Education	amenity	college, kindergarden, library, school, university
Healthcare	amenity	clinic, doctors, hospital, nursing_home, pharmacy, social_facility
Cultural/social gathering	amenity	community_centre, social_centre, nightclub
Facilities	amenity	give_box, shower
Other	amenity	internet_cafe, kitchen, marketplace, place_of_worship, public_bath, refugee_site
Natural features, land use and other infrastructure		
Open places	leisure	park, playground
Roads for motor vehicle traffic	highway	motorway, trunk, primary, secondary, tertiary, residential
Roads for primarily pedestrian traffic	highway	living_street, pedestrian, footway, bridleway, steps, path, sidewalk, crossing
Natural barriers	waterway	river
Land use	landuse	industrial, residential, retail, military, landfill

4 Methods

The analysis proceeds in two main steps. First, a multilevel confirmatory factor analysis (MCFA) model is estimated according to the specification in Table 2, using NUTS region as cluster variable. Factor scores are then computed at the regional level for each factor at regional level. These scores provide aggregate measures of social cohesion and serve as direct estimators in the second step. In this step, multivariate small-area models are estimated in three variants: (M1) without any auxiliary data, (M2) with administrative data, and (M3) with administrative and OSM data. All small area estimates are validated against figures reported in the existing literature, and improvements in estimator efficiency are assessed by comparing the Mean Squared Error (MSE) across the models.

4.1 Multilevel Confirmatory Factor Analysis

The two-level MCFA model can be defined in terms of its within- and between-level components. At the within-level, the model is:

$$y_{ijk} = \alpha_{jk} + \lambda_{Wk}\eta_{Wij} + \epsilon_{Wijk}, \tag{1}$$

where y_{ijk} denotes the observed value of respondent $i = 1, ..., I$ in region $j = 1, ..., J$ on indicator $k = 1, ..., K$. Here, α_{jk} is the intercept of indicator k in region j, λ_{Wk} is the within-level factor loading λ_W of indicator variable k. η_{Wij} is the score of the respondent i in region j on the within-level latent factor, and ϵ_{Wijk} is the within-level residual. At the between-level, the model is specified as:

$$\alpha_{jk} = \nu_k + \lambda_{Bk}\eta_{Bj} + \epsilon_{Bjk}, \tag{2}$$

where ν_k denotes the grand (cross-region) intercept of indicator k, λ_{Bk} is the between-level factor loading of indicator variable k, η_{Bj} is the score of region j on the between-level latent factor, and ϵ_{Bjk} is the between-level error term, also referred to as random intercept in multilevel analysis. The within and between components of the model are connected via the intercept a_{jk}, which acts as a latent random variable. Its variability is explained by the between-level factor η_{Bj}, while any remaining variance is captured by the between–level residual ϵ_{Bjk}.

In matrix notation, MCFA can be expressed through its variance structure. The total covariance matrix of the observed indicators is decomposed into within- and between-level components:

$$V(y_{gi}) = \Sigma_B + \Sigma_W, \tag{3}$$
$$\Sigma_B = \Lambda_B \Psi_B \Lambda_B^T + \Theta_B, \tag{4}$$
$$\Sigma_W = \Lambda_W \Psi_W \Lambda_W^T + \Theta_W, \tag{5}$$

where Σ_W and Σ_B are the modeled covariance matrices, Λ_W and Λ_B are the matrices of factor loadings, Ψ_W and Ψ_B are the factor covariance matrices, while Θ_W and Θ_B the covariance matrices of the residuals.

For the present model, all indicators are treated as continuous and residual covariance matrices Θ are estimated directly from the data.

Estimation of factor scores

Once a measurement model is determined, factor scores can be computed for both within- and between-level latent factors. For within-level factors with continuous indicators, factor scores can be estimated using the regression method (Thomson (1935); Thurstone (1935)), which corresponds to the posterior mean of the latent factor given the observed data. The factor scores for an individual i are given by:

$$\hat{F}_{\eta_i} = \hat{\Psi}\hat{\Lambda}^T(\Sigma(\hat{\Theta}))^{-1} y_i^T, \tag{6}$$

where $\hat{\Lambda}$ is the estimated matrix of factor loadings, $\hat{\Psi}$ is the estimated covariance matrix of latent factors, $\Sigma(\hat{\Theta})$ is the estimated covariance matrix of the residuals of observed responses, and y_i is the vector of observed responses for individual i.

In MCFA, cluster-level indicators correspond to random intercepts, which are latent variables themselves. Between-level factor scores can be estimated using empirical Bayes methods in combination with full-information ML estimation, with posterior means computed conditional on the model and observed data (Asparouhov and Muthén (2015)). The estimation of the model and the factor scores is performed in *Mplus*, Version 8.11.

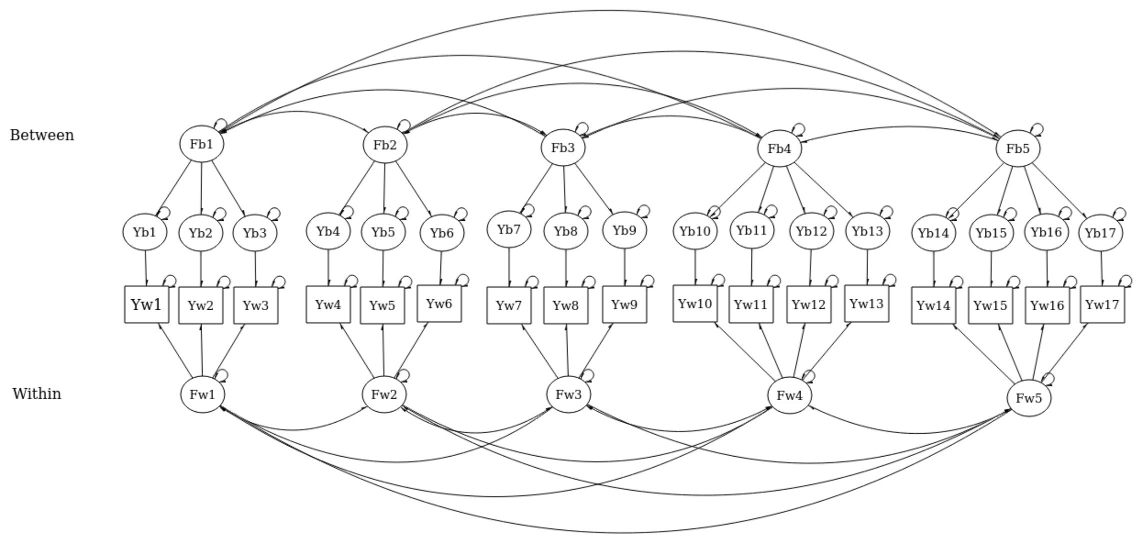

Figure 1: Latent model of social cohesion

4.2 Small Area Estimation approach

Consider a finite population U with size N divided into D small areas (domains) with indexes $d = 1, ..., D$ and size N_d. The sample size n is the sum $\Sigma_d n_d$ of small area sample sizes n_d. Let $y_d = (y_{d1}, ..., y_{dK})^T$ denote a vector of direct survey estimators for Y_d based on a random sample s of size n selected from the population U. Y_d denotes the population mean of K characteristics of interest in small area d.

The univariate Fay and Herriot (1979) model, with $K = 1$, is defined in two stages. The first stage consists of the sampling model:

$$y_d = Y_d + e_d, e_d \sim N(0, \sigma_d^2), \tag{7}$$

where e_d represents the sampling error, assumed to be independent across and within small areas, with known variance σ_d^2. The sampling model accounts for the sampling variability of the direct estimators y_d, whereas in the second stage, the population means, Y_d are linked to the auxiliary variables via the linking model:

$$Y_d = X_d\beta + u_d, u_d \sim N(0, \sigma_u^2), \tag{8}$$

where $X_d = diag(x_{d1}, ..., x_{dK})_{K\times p}$ is a matrix of auxiliary variables for small area d, $\beta = (\beta_1^T, ..., \beta_K^T)_{p\times 1}^T$ is a vector of fixed effect coefficients, and u_d is a random effect capturing unexplained variability across small areas. The univariate model is extended to the multivariate model by allowing multiple characteristics K per small area d. In a multivarite model, sampling errors $e_d \sim N(0, V_{e_d})$ are independent across small areas with known $K \times K$ covariance matrices V_{e_d} and the $K \times K$ covariance matrix V_{u_d} of random effects depends on m unknown parameters $\Theta = (\theta_1, ..., \theta_m)$. The multi-variate Fay-Herriot model proposed by Benavent and Morales (2016) is thus defined as:

$$y_d = X_d\beta + u_d + e_d, e_d \sim N(0, V_{e_d}), u_d \sim N(0, V_{u_d}), d = 1, ..., D. \tag{9}$$

In matrix form, the multivariate Fay-Herriot model is:

$$y = X\beta + Zu + e, \tag{10}$$

where y is a $DK \times 1$ vector of direct estimates, X is a $DK \times p$ matrix of auxiliary variables, Z is a $DK \times DK$ diagonal matrix, whose d^{th} column is an indicator taking the value 1 if the unit belongs to area d and 0 otherwise. Furthermore, u denotes a $DK \times 1$ vector of random effects for area d and e is a $DK \times 1$ vector of sampling errors. The Multivariate Empirical Best Linear Unbiased Predictor (MEBLUP) of Y by the use of Restricted Maximum Likelihood (REML) method is given by:

$$\hat{Y}^{MEBLUP} = X\hat{\beta} + Z\hat{u}, \tag{11}$$

$\hat{\beta}$ being the estimator of β, and $\hat{u}$ the estimator of u. MSE is estimated as a measure of uncertainty of the $\hat{Y}^{MEBLUP}$ using the method proposed by Benavent and Morales (2016).

In this thesis, small areas are defined as NUTS-3 regions. Although the model is specified at the NUTS-3 level, for countries where NUTS-3 regions are not available, values are taken from the highest available level (NUTS-2 or NUTS-1) and are assumed to be homogeneous across all NUTS-3 regions within that higher-level unit. Small area estimates for NUTS-3 regions in these countries will rely entirely on auxiliary data.

The small area model and MSE are both estimated using the R package *msae*, Version 0.1.5.

References

Asparouhov, Tihomir, and Bengt Muthén. 2015. "General Random Effect Latent Variable Modeling." In, 163–92. Emerald Publishing Limited. https://doi.org/10.1108/978-1-68123-329-120251009.

Balducci, Francesco. 2019. "Is OpenStreetMap a Good Source of Information for Cultural Statistics? The Case of Italian Museums." *Environment and Planning B: Urban Analytics and City Science* 48 (3): 503–20. https://doi.org/10.1177/2399808319876949.

Barrington-Leigh, Christopher, and Adam Millard-Ball. 2017. "The World's User-Generated Road Map Is More Than 80." Edited by Mohammad Ali. *PLOS ONE* 12 (8): e0180698. https://doi.org/10.1371/journal.pone.0180698.

Basu, Nandita, Oscar Oviedo-Trespalacios, Mark King, Md. Kamruzzaman, and Md. Mazharul Haque. 2022. "The Influence of the Built Environment on Pedestrians' Perceptions of Attractiveness, Safety and Security." *Transportation Research Part F: Traffic Psychology and Behaviour* 87 (May): 203–18. https://doi.org/10.1016/j.trf.2022.03.006.

Benavent, Roberto, and Domingo Morales. 2016. "Multivariate Fay–Herriot Models for Small Area Estimation." *Computational Statistics & Data Analysis* 94 (February): 372–90. https://doi.org/10.1016/j.csda.2015.07.013.

Berger-Schmitt, Regina. n.d. "Considering Social Cohesion in Quality of Life Assessments: Concept and Measurement." In, 403–28. Kluwer Academic Publishers. https://doi.org/10.1007/0-306-47513-8_18.

Bollen, Kenneth A., and Rick H. Hoyle. 1990. "Perceived Cohesion: A Conceptual and Empirical Examination." *Social Forces* 69 (2): 479. https://doi.org/10.2307/2579670.

Bottoni, Gianmaria. 2016. "A Multilevel Measurement Model of Social Cohesion." *Social Indicators Research* 136 (3): 835–57. https://doi.org/10.1007/s11205-016-1470-7.

Brakel, Jan A van den, Harm Jan Boonstra, and Marc Smeets. 2025. *Multilevel Time Series Models for Small Area Estimation of Social Cohesion Indicators Based on the Dutch Social Cohesion and Well-Being Survey*. Statistics Netherlands.

Chan, Joseph, Ho-Pong To, and Elaine Chan. 2006. "Reconsidering Social Cohesion: Developing a Definition and Analytical Framework for Empirical Research." *Social Indicators Research* 75 (2): 273–302. https://doi.org/10.1007/s11205-005-2118-1.

Cho, Gihyoug, Daniel A. Rodríguez, and Asad J. Khattak. 2009. "The Role of the Built Environment in Explaining Relationships Between Perceived and Actual Pedestrian and Bicyclist Safety." *Accident Analysis & Prevention* 41 (4): 692–702. https://doi.org/10.1016/j.aap.2009.03.008.

Christie, Richard, and Gordon W. Allport. 1954. "The Nature of Prejudice." *The American Journal of Psychology* 67 (4): 742. https://doi.org/10.2307/1418507.

Czaika, Mathias, and Armando Di Lillo. 2018. "The Geography of Anti-Immigrant Attitudes Across Europe, 2002–2014." *Journal of Ethnic and Migration Studies* 44 (15): 2453–79. https://doi.org/10.1080/1369183x.2018.1427564.

Duhaime, Gérard, Edmund Searles, Peter J. Usher, Heather Myers, and Pierre Fréchette. 2004. "Social Cohesion and Living Conditions in the Canadian Arctic: From Theory to

Measurement." *Social Indicators Research* 66 (3): 295–318. https://doi.org/10.1023/b:soci.0000003726.35478.fc.

Durkheim, Emile. 2018. "The Division of Labor in Society (1983)." In, 217222. Routledge.

Dustmann, Christian, and Anna Okatenko. 2014. "Out-Migration, Wealth Constraints, and the Quality of Local Amenities." *Journal of Development Economics* 110 (September): 52–63. https://doi.org/10.1016/j.jdeveco.2014.05.008.

Fay, Robert E., and Roger A. Herriot. 1979. "Estimates of Income for Small Places: An Application of James-Stein Procedures to Census Data." *Journal of the American Statistical Association* 74 (366a): 269–77. https://doi.org/10.1080/01621459.1979.10482505.

Feld, Scott L. 1981. "The Focused Organization of Social Ties." *American Journal of Sociology* 86 (5): 1015–35. https://doi.org/10.1086/227352.

Ferdinand, TÃ et al. 1887. *Community and Society*. Transaction Publishers.

Fitzgerald, Jennifer, and Jennifer Wolak. 2014. "The Roots of Trust in Local Government in Western Europe." *International Political Science Review* 37 (1): 130–46. https://doi.org/10.1177/0192512114545119.

Foster, Sarah, Lisa Wood, Hayley Christian, Matthew Knuiman, and Billie Giles-Corti. 2013. "Planning Safer Suburbs: Do Changes in the Built Environment Influence Residents' Perceptions of Crime Risk?" *Social Science & Medicine* 97 (November): 87–94. https://doi.org/10.1016/j.socscimed.2013.08.010.

Fraser, Timothy, Osama Awadalla, Harshita Sarup, and Daniel P. Aldrich. 2024. "A Tale of Many Cities: Mapping Social Infrastructure and Social Capital Across the United States." *Computers, Environment and Urban Systems* 114 (December): 102195. https://doi.org/10.1016/j.compenvurbsys.2024.102195.

Fraser, Timothy, Takahiro Yabe, Daniel P. Aldrich, and Esteban Moro. 2024. "The Great Equalizer? Mixed Effects of Social Infrastructure on Diverse Encounters in Cities." *Computers, Environment and Urban Systems* 113 (October): 102173. https://doi.org/10.1016/j.compenvurbsys.2024.102173.

Goodchild, Michael F. 2013. "The Quality of Big (Geo)data." *Dialogues in Human Geography* 3 (3): 280–84. https://doi.org/10.1177/2043820613513392.

Heine, Cate, Timur Abbiasov, Paolo Santi, and Carlo Ratti. 2025. "The Role of Urban Amenities in Facilitating Social Mixing: Evidence from Stockholm." *Landscape and Urban Planning* 254 (February): 105250. https://doi.org/10.1016/j.landurbplan.2024.105250.

Herfort, Benjamin, Sven Lautenbach, João Porto de Albuquerque, Jennings Anderson, and Alexander Zipf. 2023. "A Spatio-Temporal Analysis Investigating Completeness and Inequalities of Global Urban Building Data in OpenStreetMap." *Nature Communications* 14 (1). https://doi.org/10.1038/s41467-023-39698-6.

Hipp, John R., Jonathan Corcoran, Rebecca Wickes, and Tiebei Li. 2014. "Examining the Social Porosity of Environmental Features on Neighborhood Sociability and Attachment." Edited by Cédric Sueur. *PLoS ONE* 9 (1): e84544. https://doi.org/10.1371/journal.pone.0084544.

Jenson, Jane. 1998. *Mapping Social Cohesion: The State of Canadian Research*. Vol. 103. Canadian policy research networks Ottawa.

Joshi, Anaya, and Daniel Aldrich. 2022. "Corralling a Chimera: A Critical Review of the

Term Social Infrastructure." http://dx.doi.org/10.32388/wq6xsb.

Maina, Joseph, Paul O. Ouma, Peter M. Macharia, Victor A. Alegana, Benard Mitto, Ibrahima Socé Fall, Abdisalan M. Noor, Robert W. Snow, and Emelda A. Okiro. 2019. "A Spatial Database of Health Facilities Managed by the Public Health Sector in Sub Saharan Africa." *Scientific Data* 6 (1). https://doi.org/10.1038/s41597-019-0142-2.

McKay, Lawrence, Will Jennings, and Gerry Stoker. 2023. "What Is the Geography of Trust? The Urban-Rural Trust Gap in Global Perspective." *Political Geography* 102 (April): 102863. https://doi.org/10.1016/j.polgeo.2023.102863.

Michelangeli, Alessandra, John Östh, Marina Toger, and Umut Türk. 2025. "Inequality in Access to Urban Amenities." *Npj Urban Sustainability* 5 (1). https://doi.org/10.1038/s42949-025-00248-2.

Mitsch, Frieder, Neil Lee, and Elizabeth Ralph Morrow. 2021. "Faith No More? The Divergence of Political Trust Between Urban and Rural Europe." *Political Geography* 89 (August): 102426. https://doi.org/10.1016/j.polgeo.2021.102426.

Moretti, Angelo, and Anisa Ahmed. 2024. "Regional Multidimensional Attitudes Towards Immigration: Evidence from the European Social Survey Using Small Area Estimation." *Social Indicators Research* 174 (1): 91–121. https://doi.org/10.1007/s11205-024-03381-0.

Ninivaggi, Federico, and Eleonora Cutrini. 2023. "Exploring Local Well-Being and Vulnerability Through Openstreetmap: The Case of Italy." *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4609064.

Noll, Heinz-Herbert. 2002. "Towards a European System of Social Indicators: Theoretical Framework and System Architecture." *Social Indicators Research* 58 (1-3): 47–87. https://doi.org/10.1023/a:1015775631413.

Orteca, Maria Katia. 2015. "An Investigation of Social Capital in Britain Using Small Area Estimation Analysis." PhD thesis, University of Essex.

Rajulton, Fernando, Zenaida R. Ravanera, and Roderic Beaujot. 2006. "Measuring Social Cohesion: An Experiment Using the Canadian National Survey of Giving, Volunteering, and Participating." *Social Indicators Research* 80 (3): 461–92. https://doi.org/10.1007/s11205-006-0011-1.

Rao, J. N. K., and Isabel Molina. 2015. "Small Area Estimation," August. https://doi.org/10.1002/9781118735855.

Salvucci, Gianluigi, and Luca Salvati. 2021. "Official Statistics, Building Censuses, and OpenStreetMap Completeness in Italy." *ISPRS International Journal of Geo-Information* 11 (1): 29. https://doi.org/10.3390/ijgi11010029.

Schiefer, David, and Jolanda van der Noll. 2016. "The Essentials of Social Cohesion: A Literature Review." *Social Indicators Research* 132 (2): 579–603. https://doi.org/10.1007/s11205-016-1314-5.

Small, Mario L., and Laura Adler. 2019. "The Role of Space in the Formation of Social Ties." *Annual Review of Sociology* 45 (1): 111–32. https://doi.org/10.1146/annurev-soc-073018-022707.

Thomson, Godfrey H. 1935. "Definition and Measurement of General Intelligence." *Nature* 135 (3413): 509–9. https://doi.org/10.1038/135509b0.

Thurstone, L. L. 1935. *The Vectors of Mind: Multiple-Factor Analysis for the Isolation of*

Primary Traits. University of Chicago Press. https://doi.org/10.1037/10018-000.

Ulrich-Schad, Jessica D. 2015. "Recreational Amenities, Rural Migration Patterns, and the Great Recession." *Population and Environment* 37 (2): 157–80. https://doi.org/10.1007/s11111-015-0238-3.

Whelan, Christopher T., and Bertrand Maître. 2005. "Economic Vulnerability, Multidimensional Deprivation and Social Cohesion in an Enlarged European Community." *International Journal of Comparative Sociology* 46 (3): 215–39. https://doi.org/10.1177/0020715205058942.

Wickes, Rebecca, Renee Zahnow, Jonathan Corcoran, and John R Hipp. 2018. "Neighbourhood Social Conduits and Resident Social Cohesion." *Urban Studies* 56 (1): 226–48. https://doi.org/10.1177/0042098018780617.

Zhou, Qi. 2017. "Exploring the Relationship Between Density and Completeness of Urban Building Data in OpenStreetMap for Quality Estimation." *International Journal of Geographical Information Science* 32 (2): 257–81. https://doi.org/10.1080/13658816.2017.1395883.